

Network Capacity Region of Multi-Queue Multi-Server Queueing System with Time Varying Connectivities

Hassaan Halabian, Ioannis Lambadaris, Chung-Horng Lung

Department of Systems and Computer Engineering

Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6 Canada

Email: {hassanh, ioannis.lambadaris, chung-horng.lung}@sce.carleton.ca

Abstract—Network capacity region of multi-queue multi-server queueing system with random ON-OFF connectivities and stationary arrival processes is derived in this paper. Specifically, the necessary and sufficient conditions for the stability of the system are derived under general arrival processes with finite first and second moments. In the case of stationary arrival processes, these conditions establish the network capacity region of the system. It is also shown that AS/LCQ (Any Server/Longest Connected Queue) policy stabilizes the system when it is stabilizable. Furthermore, an upper bound for the average queue occupancy is derived for this policy.

I. INTRODUCTION

Resource allocation is one of the main concerns in the design process of emerging wireless networks. Examples of such networks are OFDMA and CDMA wireless systems in which orthogonal resources (OFDM subcarriers and CDMA codes) must be allocated to multiple users. Research in this area focuses on finding optimal policies to allocate orthogonal subchannels to the users. There are stochastic arrivals for each user which may be buffered to be transmitted in the future. Therefore, the resource allocation problem can be modeled as a multi-queue multi-server queueing system with parallel queues competing for available servers (which may model orthogonal subchannels [1], [2], [3], [4], [11], [15]). However, because of users mobility, environmental changes, fading and etc., connectivity of each queue to each server is changing with time randomly. Thus, we are faced to a multi-queue multi-server system with time varying channel quality for which we have to design an appropriate server allocation policy. One of the main performance attributes which must be considered for each policy is its capacity region and how much this region coincides with the *network capacity region* [8]. The capacity region of a network is defined as the closure of the set of all arrival rate matrices for which there exists an appropriate policy that stabilizes the system [8]. This region is unique for each network and is independent of resource allocation policy. On the other hand, the capacity region of a specified policy, say π , is the closure of the set of all arrival rate matrices for which π results into the stability of the system. Obviously,

the capacity region of any policy is a subset of the network capacity region. In fact, the network capacity region of a system is the union of the capacity regions of all the possible resource allocation policies we can have for a network [8]. A policy that achieves the network capacity region is called *throughput optimal*.

The stability problem in wireless queueing networks was mainly addressed in [6], [7], [8], [9]. In [6], authors introduced the capacity region of a queueing network. They considered a time slotted system in their work and assumed that arrival processes are i.i.d. sequences and the queue length process is a Markov process. They also characterized the network capacity region of multi-queue single-server system with time varying ON-OFF connectivities which is described by some conditions on the arrival traffic [7]. They also proved that for a symmetric system (with the same arrival and connectivity statistics for all the queues), LCQ (Longest Connected Queue) policy maximizes the capacity region and also provides the optimal performance in terms of average queue occupancy (or equivalently average delay) [7]. In [8], [9] and [13], the notion of network capacity region of a wireless network was introduced for more general arrival and queue length processes. Furthermore, Lyapunov drift techniques were applied in [8] and [9] to analyse the stability of the proposed policies for stochastic optimization problems in wireless networks.

The problem of server allocation in multi-queue multi-server systems with time varying connectivities was mainly addressed in [1], [2], [3], [4], [11]. In [2], Maximum Weight (MW) policy, a throughput optimal server allocation policy for stationary connectivity processes was proposed. However, [2] does not explicitly mention the conditions on the arrival traffic to guarantee the stability of MW. References [1], [3], [4], [11] study the optimal server allocation problem in terms of average delay. In [1], [3], [4], authors argue that in general, achieving instantaneous throughput and load balancing is impossible in a policy. However, as they show this goal is attainable in the special case of ON-OFF connectivity processes. They also introduced the MTLB (Maximum-Throughput Load-Balancing) policy and showed that this policy is minimizing a class of cost functions including total average delay for the case of two symmetric queues (with the same arrival and connectivities statistics). [11] considers this problem for general number of symmetric queues and servers. Authors in [11] characterized

¹This work was supported by Mathematics of Information Technology and Complex Systems (MITACS) and Natural Sciences and Engineering Research Council of Canada (NSERC).

a class of *Most Balancing* (MB) policies among all work conserving policies which are minimizing a class of cost functions including total average delay in stochastic ordering sense. They used stochastic ordering and dynamic coupling arguments to show the optimality of MB policies for symmetric systems.

In this paper, we will characterize the capacity region of multi-queue multi-server queueing system with random ON-OFF connectivities and stationary arrival processes based on the stochastic properties of the system. Toward this, the necessary and sufficient conditions for the stability of the system is derived under a general arrival process with finite first and second moments. For stationary arrival processes, these conditions establish the network capacity region of the system. We also showed that a simple server allocation policy called AS/LCQ maximizes the capacity region i.e. its capacity region coincide with the network capacity region and therefore it is a throughput optimal policy. It is worth mentioning that AS/LCQ acts exactly the same as MW policy proposed in [2] when the connectivity process is ON-OFF in MW.

The rest of the paper is organized as follows. Section II describes the model and notation required through the paper. In section III we discuss about the *strong stability* definition in queueing networks and Lyapunov drift technique briefly. Then, we will derive necessary and sufficient conditions for the stability of our model and also find an upper bound for the average queue occupancy. In section IV we present simulation results and compare stability and delay performances of some heuristic work-conserving policies with those of AS/LCQ and the upper bound obtained in section III. Section V summarizes the conclusions of the paper.

II. MODEL DESCRIPTION

Our model in this paper is the same as the model used in [1], [2], [3], [4], [11] with ON-OFF connectivity processes. We consider a time slotted queueing system with equal length time slots and equal length packets. The model consists of a set of parallel queues \mathcal{L} and a set of identical servers \mathcal{K} . Each server can serve at most one packet at each time slot and we do not allow server sharing by the queues. In other words, each server can serve at most one queue at each time slot. Assume that $|\mathcal{L}| = L$ and $|\mathcal{K}| = K$. At each time slot t , the link between each queue $i \in \{1, \dots, L\}$ and server $j \in \{1, \dots, K\}$ is either connected or disconnected. Assume that connectivity process between queue i and server j is modelled by an i.i.d. binary random process which is denoted by $G_{ij}(t)$, i.e. $G_{ij}(t) \in \{0, 1\}$. Suppose that p_{ij} represents the expected value of this process, i.e. $E[G_{ij}(t)] = p_{ij}$. There are also exogenous arrival processes to the queues in set \mathcal{L} . Assume that the arrival process to each queue i at time slot t (i.e. the number of packet arrivals during time slot t) is represented by $A_i(t)$. For these processes we assume that $E[A_i^2(t)] < A_{max}^2 < \infty$ for all t . Each queue has an infinite buffer space i.e. we do not have packet drops. We assume that the new arrivals are added to each queue at the end of each time slot. Let $X(t) = (X_1(t), \dots, X_L(t))$ be the queue length process vector at the end of time slot t after adding new arrivals to the queues. Figure 1 shows the model used in this paper.

A server scheduling policy at each time slot should decide on how to allocate servers from set \mathcal{K} to the queues in set \mathcal{L} . This must be accomplished based on the available information about the connectivities $G_{ij}(t)$ and also the queue length process $X(t)$.

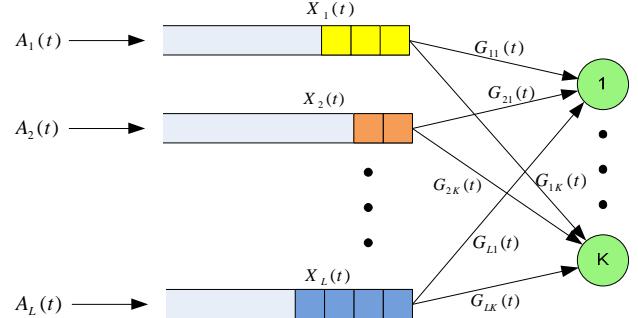


Fig. 1: Multi-queue multi-server queueing system with time varying connectivities

III. STABILITY OF MULTI-QUEUE MULTI-SERVER SYSTEM WITH TIME VARYING CONNECTIVITIES

In this section, we will consider the stability problem of multi-queue multi-server system with ON-OFF connectivities for which we will find the necessary and sufficient conditions for its stability. We also show that AS/LCQ (Any Server/Longest Connected Queue) policy will stabilize the system as long as it is stabilizable. The details of this policy will be presented in part D of this section. At first, we will have a review on the notion of strong stability in queueing networks.

A. Strong Stability

We begin with introducing the definition of strong stability for a queueing system [8], [9]. Other definitions can be found in [5], [6], [7], [14]. Consider a discrete time single queue system with an arrival process $A(t)$ and service process $\mu(t)$. Assume that the arrivals are added to the system at the end of each time slot. We can see that the queue length process $X(t)$ at time t evolves with time according to the following rule.

$$X(t) = (X(t-1) - \mu(t))^+ + A(t) \quad (1)$$

where $(\cdot)^+$ outputs the term inside the brackets if it is nonnegative and is zero otherwise. Strong stability is given by the following definition [8].

Definition 1: A queue satisfying the conditions above is called *strongly stable* if

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} E[X(\tau)] < \infty \quad (2)$$

Naturally for a queueing system we have the following definition [8].

Definition 2: A queueing system is called to be strongly stable if all the queues in the system are strongly stable.

In our work we use the strong stability definition and from now we use “stability” and “strong stability” interchangably.

The following important property of strongly stable queues gives an invaluable insight of the above definitions.

Lemma 1 [8]: If a queue is strongly stable and either $E[A(t)] \leq A$ for all t or $E[\mu(t) - A(t)] \leq D$ where A and D are finite nonnegative constants, then

$$\lim_{t \rightarrow \infty} \frac{1}{t} E[X(t)] = 0 \quad (3)$$

A very important and useful mathematical tool used in network stability analysis and stochastic control/optimization of wireless networks is *Lyapunov Drift* technique. We now present a brief review of this technique.

B. Lyapunov Drift

The basic idea behind the Lyapunov stability method is to define a nonnegative function of queue backlog in a queueing system which can be seen as a *measure* of the total aggregated backlog in the system at time t . Then we evaluate the “drift” of such function in two successive time slots by taking the effect of our control decision (scheduling or resource allocation policy) into account. If the expected value of the drift is negative as the backlog goes beyond a fixed threshold, then the system is stable. This is the method used in [8], [9], [12], [10], [13], [14] to prove the stability of the systems working under their proposed policies.

For a queueing system with L queues and queue length vector $X(t) = (X_1(t), \dots, X_L(t))$, the following quadratic Lyapunov function has been used in literature ([8], [9], [12], [13], [14]).

$$V(X) = \sum_{i=1}^L X_i^2(t) \quad (4)$$

Assume that $E[X_i(0)] < \infty, \forall i = 1, 2, \dots, L$ and $X(t)$ evolves with some probabilistic law (not necessarily Markovian). Then, the following important lemma holds.

Lemma 2 [8]: If there exist constants $B > 0$ and $\epsilon > 0$ such that for all time slots t we have

$$E[V(X(t+1)) - V(X(t)) \mid X(t)] \leq B - \epsilon \sum_{i=1}^L X_i(t), \quad (5)$$

then the system is strongly stable and further we have

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{i=1}^L E[X_i(\tau)] \leq \frac{B}{\epsilon} \quad (6)$$

The left hand side of expression (5) is usually called Lyapunov drift function which is a measure of expected value of changes in the backlog in two successive time slots. We can easily see the idea behind Lyapunov method in stabilizing queueing systems from Lemma 1. It is not hard to show that, when the aggregated backlog in the system goes beyond the bound $\frac{B}{\epsilon}$, then the Lyapunov drift in the left hand side of (5) will be negative, meaning that the system receives a negative drift on the expected aggregated backlog in two successive time slots. In other words the system tends toward lower backlog and this results in its stability.

C. Necessary Condition for the Stability of the System

Let $h_{ik}(t)$ be the departure process at time slot t from queue i to server k . Then, we can have the following equation for the queue length process which shows the evolution of queue length process with time.

$$X_i(t) = X_i(t-1) - \sum_{k=1}^K h_{ik}(t) + A_i(t) \quad (7)$$

To find the necessary condition for the stability of the system, we need to use the following lemma.

Lemma 3: If the system is strongly stable under some server allocation policy π , then for each queue i

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t E[A_i(\tau)] = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_{k=1}^K E[h_{ik}(\tau)], \quad (8)$$

i.e. for a stable system the average expected arrivals to a queue is equal to the average expected departure from that queue.

Proof: See appendix A.

We now proceed to find the necessary condition for the stability of the system.

Theorem 1: If there exists a server allocation policy π under which the system is stable, then

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_{i \in Q} E[A_i(\tau)] \leq K - \sum_{k=1}^K \prod_{i \in Q} (1 - p_{ik}) \quad (9)$$

$$\forall Q \subset \{1, \dots, L\}$$

Proof: See appendix B.

Remark: If the arrival processes $A_i(t)$ ’s are stationary, then $E[A_i(t)] = \lambda_i$ for all t and therefore the left hand side of (9) will be equal to $\sum_{i \in Q} \lambda_i$. Consequently, the necessary condition for the stability of the system with stationary arrival processes would be

$$\sum_{i \in Q} \lambda_i \leq \sum_{k=1}^K \left(1 - \prod_{i \in Q} (1 - p_{ik})\right) \quad \forall Q \subset \{1, \dots, L\}. \quad (10)$$

D. Sufficient Condition for the Stability of the System

We can divide the server allocation policy in our model into two scheduling problems. First, we should determine the order under which servers are selected for service and second, for each server decide to allocate it to a particular queue. Consider the policy that chooses an arbitrary ordering of servers and then for each server, allocates it to its longest connected queue (LCQ). In other words, in this policy we do not restrict ourselves with a specific ordering of servers and we accept any permutation of the servers according to which servers will be selected for service. However, for the next phase of scheduling, for each selected server we use the LCQ policy. We call such a policy as AS/LCQ (Any Server/Longest Connected Queue). We will now derive the sufficient condition for the stability of our model and prove that AS/LCQ stabilizes the system as long as condition (11) is satisfied. An upper bound is also derived for the time averaged expected number of packets in the system.

Theorem 2: The multi-queue multi-server system is stable under AS/LCQ if for all t

$$\sum_{i \in Q} E[A_i(t)] < K - \sum_{k=1}^K \prod_{i \in Q} (1 - p_{ik}) \quad \forall Q \subset \{1, \dots, L\}. \quad (11)$$

Furthermore, the following bound for the average expected “aggregate” occupancy holds.

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{i=1}^L E[X_i(\tau)] \leq \frac{-\frac{L}{2} (LA_{max}^2 + K(2K-1))}{\max_{Q \subset \{1, 2, \dots, L\}, t} \left\{ \sum_{i \in Q} E[A_i(t)] - K + \sum_{k=1}^K \prod_{i \in Q} (1 - p_{ik}) \right\}} \quad (12)$$

Proof: See appendix C.

It is worth mentioning that AS/LCQ acts exactly the same as MW policy proposed in [2] when the connectivity process is ON-OFF in MW.

Note that for all the servers we only use the backlog information at the beginning of each time slot, i.e. during the implementation of AS/LCQ policy at each time slot we do not update the queue lengths until all the servers are allocated at which point we update the queue lengths. It is interesting to note that this policy can be non-work conserving at some time slots. In other words, there may exist some idle servers at a time slot while they could have served other backlogged queues. We will discuss about it through an example in part E of this section.

Remark: Note that by considering stationary assumption on the arrival processes, the condition (11) would be

$$\sum_{i \in Q} \lambda_i < K - \sum_{k=1}^K \prod_{i \in Q} (1 - p_{ik}) \quad \forall Q \subset \{1, \dots, L\}. \quad (13)$$

According to the definition of system capacity region and (10) and (13), equation (10) characterizes the network capacity region of multi-queue multi-server system with stationary ON-OFF connectivities and stationary arrivals.

E. Discussion

As mentioned earlier, AS/LCQ may exhibit non-work conserving behavior during some time slots. This can be clarified by the following example.

Consider a system with $L = 2$ and $K = 3$ with queue length vector $X(t) = (2, 1)$ at time slot t . For the connectivities at this time we have the following matrix.

$$G(t) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Assume that the ordering of server selection is server 1 first and then server 2 and finally server 3. Servers 1 and 2 both are allocated to queue 1 according to LCQ rule. Server 3 is connected to both of the queues. Since in AS/LCQ all the servers are allocated first and then the queue lengths are updated afterwards, queue 1 is the longest connected queue for server 3. Thus, server 3 is allocated to queue 1 as well.

However, queue 1 has only two packets waiting for service and therefore server 3 will be idle at this time slot (although it could have been used to serve queue 2).

Note that since AS/LCQ is a non-work conserving policy it can not be delay optimal. However, it can achieve the network capacity region as explained previously in part D. In fact, AS/LCQ will exhibit non-work conserving behaviour in light arrival loads and as the load increases its behaviour will converge to work conserving. Since the capacity region of a system is mainly determined by its behaviour in heavy arrival loads, this property of AS/LCQ does not have conflict with its throughput optimality. It is worth mentioning that not all work-conserving policies are throughput optimal. In the following section by simulations we will observe that some work-conserving policies cannot achieve the network capacity region. In the following section, we will also observe that how the service ordering of servers affects the average total queue occupancy. However, as we showed in the previous part an arbitrary ordering is sufficient to achieve the network capacity region.

IV. SIMULATION RESULTS

Simulation is used to show the validity of our analysis in the previous section and also to compare performance of AS/LCQ to some heuristic work conserving policies including LCSF/LCQ (Least Connected Server First/Longest Connected Queue), MCSF/LCQ (Most Connected Server First/Longest Connected Queue), LCSF/SCQ (Least Connected Server First/Shortest Connected Queue), MCSF/SCQ (Most Connected Server First/Shortest Connected Queue) and a Randomized policy [11].

The LCSF (MCSF) policy at the first phase of scheduling (i.e. determination of servers order) will sort the servers for service according to their number of connectivities in an ascending (descending) order. The LCQ (SCQ) policy will assign the selected server it to its longest connected queue (shortest connected queue). Note that in order to make SCQ policies work conserving, we only serve the shortest non-empty queues. The Randomized policy at each time slot makes random server selections and for each server random non-empty queue selection.

We have simulated a system consisting of 16 queues ($L = 16$) and 4 servers ($K = 4$). First, we considered a symmetric system in which all the arrivals to all the queues are the same in distribution. We also assumed that connectivity variables have the same distribution (the same connectivity probabilities). In this system, arrivals are assumed to have i.i.d. Bernoulli distributions. The capacity region for these special cases would be an n dimensional cube whose side size is equal to $\frac{K-K(1-p)^L}{L}$ which is 0.243 for $p = 0.2$ and is almost 0.25 for $p = 0.9$. Figures (2) and (3) show the average total occupancy of different policies for connectivity probabilities 0.2, and 0.9 versus arrival rate per queue. In these figures, it is observed that in all the cases if the arrivals are inside the capacity region, AS/LCQ can stabilize the system and has average total occupancy below the bound we derived in the previous section.

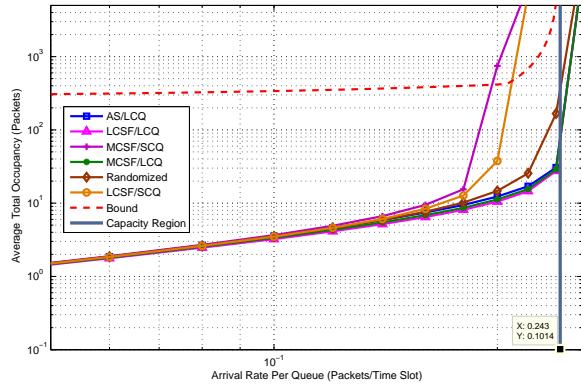


Fig. 2: Average Total Occupancy for $p = 0.2$

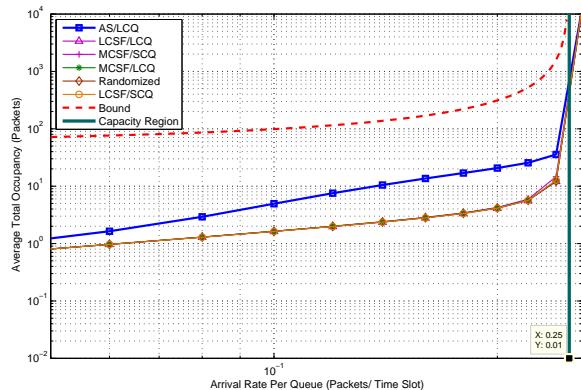


Fig. 3: Average Total Occupancy for $p = 0.9$

We can further conclude that as the connectivity variable increases the performance of the work conserving policies become the same. This agrees with intuition since when the system is close to full connectivity, any work conserving algorithm will be optimal in terms of average occupancy and of course better than any non-work conserving policy like AS/LCQ. Although AS/LCQ has larger average total occupancy compared to other policies, it still stabilizes the system as long as arrivals are inside the capacity region and has bounded average total occupancy. However, this is not the case for LCSF/SCQ and MCSF/SCQ policies and they cannot stabilize the system for certain arrivals inside the capacity region. From these figures we also see that randomized policy performs very close to the other policies in these special cases and this is due to existence of symmetry (in arrivals and connectivities) in these cases.

We have also simulated an asymmetric system in which connectivity variables comes from the following matrix in which $p_{ij} = E[G_{ij}(t)]$. This matrix was chosen randomly.

$$p = \begin{pmatrix} 0.9 & 0.2 & 0.2 & 0.8 & 0.2 & 0.1 & 0.5 & 0.6 \\ 0.8 & 0.1 & 0.1 & 0.9 & 0.02 & 0.5 & 0.8 & 0.8 \\ 0.9 & 0.02 & 0.5 & 0.99 & 0.3 & 0.8 & 0.78 & 0.99 \\ 0.8 & 0.03 & 0.9 & 0.87 & 0.5 & 0.98 & 0.62 & 0.4 \end{pmatrix}$$

$$\begin{pmatrix} 0.2 & 0.72 & 0.86 & 0.3 & 0.66 & 0.21 & 0.84 & 0.03 \\ 0.1 & 0.65 & 0.65 & 0.15 & 0.58 & 0.32 & 0.69 & 0.12 \\ 0.02 & 0.42 & 0.94 & 0.35 & 0.9 & 0.16 & 0.96 & 0.21 \\ 0.8 & 1 & 0.7 & 0.09 & 0.1 & 0.45 & 0.13 & 0.07 \end{pmatrix}^T$$

In this experiment, arrivals are following the Poisson distribution with the same rates for each slot and each queue. The capacity region in this case is not easy to characterize and describe in a concise manner. Figure (4) shows the results for this case. In this figure, we observe that Randomized policy could not capture the capacity region wholly. However, LCQ policies (AS/LCQ, LCSF/LCQ and MCSF/LCQ) performs similarly to each other from stability point of view.

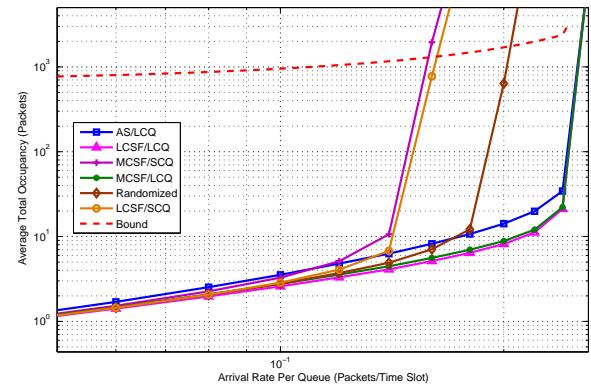


Fig. 4: Average total occupancy for an asymmetric scenario

From the above simulations, we can also observe that AS/LCQ performs slightly worse as compared with other policies in light arrival loads. This behaviour is because the fact that AS/LCQ may exhibit non-work conserving behaviour more frequently for light arrival loads. However, as the load increases AS/LCQ will be work conserving with high probability. We can also observe that the obtained bound is not tight.

V. CONCLUSIONS

In this paper we derived the necessary and sufficient conditions for the stability of multi-queue multi-server system with random connectivities and characterized the capacity region of this system for stationary arrivals. We also introduced AS/LCQ policy and argued that although this policy is a non-work conserving policy, it can stabilize the system for all the arrivals inside the capacity region and therefore it is a throughput optimal policy. Then, we derived an upper bound of the average queue occupancy for this policy. Finally, we used simulations to validate our analysis and compare this policy to some work conserving policies in terms of average queue occupancy.

If we modify the policy such that the queue lengths are updated after each server is allocated, we can establish a work conserving policy. However, this does not increase the capacity region for a system with stationary arrivals. However, we may help us to obtain a tighter bound than we obtained in this work.

APPENDIX A
PROOF OF LEMMA 3

Proof: If we write equation (7) for $\tau = 1, 2, \dots, t$ and then adding them up, we will have

$$X_i(t) = X_i(0) - \sum_{\tau=1}^t \sum_{k=1}^K h_{ik}(\tau) + \sum_{\tau=1}^t A_i(\tau) \quad (14)$$

Taking the expectation from both sides, dividing by t and then taking the limit as t goes to infinity, we will have the following.

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{E[X_i(t)]}{t} &= \lim_{t \rightarrow \infty} \frac{E[X_i(0)]}{t} \\ - \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_{k=1}^K E[h_{ik}(\tau)] &+ \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t E[A_i(\tau)] \end{aligned} \quad (15)$$

According to Lemma 1 and the assumption that $E[X_i(0)] < \infty$, the left hand side term and the first term in the right hand side term are equal to zero and therefore the result is proven. \square

APPENDIX B
PROOF OF THEOREM 1

Since the system is strongly stable, (8) must be satisfied for any subset of queues $Q \subset \{1, \dots, L\}$, i.e.

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_{i \in Q} E[A_i(\tau)] = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_{i \in Q} \sum_{k=1}^K E[h_{ik}(\tau)] \quad (16)$$

We now define the sets $B_k(\tau)$ as

$$B_k(\tau) = \{G_{\ell k}(\tau), X_{\ell}(\tau-1), \ell \in Q\} \quad (17)$$

For $B_k(\tau)$, three disjoint cases are imaginable.

$$\begin{aligned} B_k^1(\tau) &= \{G_{ik}(\tau) = 0, i \in Q\} \\ B_k^2(\tau) &= \{G_{ik}(\tau) = 0, i \in Q\}^c \cap \{X_i(\tau-1) = 0, i \in Q\} \\ B_k^3(\tau) &= \{G_{ik}(\tau) = 0, i \in Q\}^c \cap \{X_i(\tau-1) = 0, i \in Q\}^c \end{aligned}$$

By conditioning each term in the right hand side summation in (16) to the event $B_k(\tau)$ we have

$$\sum_{k=1}^K \sum_{i \in Q} E[h_{ik}(\tau)] = \sum_{k=1}^K E_{B_k(\tau)} \left[E \left[\sum_{i \in Q} h_{ik}(\tau) \mid B_k(\tau) \right] \right] \quad (18)$$

We can easily see that

$$E \left[\sum_{i \in Q} h_{ik}(\tau) \mid B_k^j(\tau) \right] = 0 \quad j = 1, 2 \quad (19)$$

and

$$E \left[\sum_{i \in Q} h_{ik}(\tau) \mid B_k^3(\tau) \right] \leq 1 \quad (20)$$

Using (19) and (20), equation (18) can be simplified to the following.

$$\sum_{k=1}^K \sum_{i \in Q} E[h_{ik}(\tau)] \leq \sum_{k=1}^K (1 - P[B_k^1(\tau)] - P[B_k^2(\tau)]) \quad (21)$$

Note that $P[B_k^2(\tau)] \geq 0$ and for $P[B_k^1(\tau)]$, we have

$$P[B_k^1(\tau)] = \prod_{i \in Q} (1 - p_{ik}) \quad (22)$$

Finally, from (16), (21) and (22) we conclude that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_{i \in Q} E[A_i(\tau)] \leq \sum_{k=1}^K (1 - \prod_{i \in Q} (1 - p_{ik})) \quad (23)$$

and the theorem follows. \square

APPENDIX C
PROOF OF THEOREM 2

Proof: We will start with the Lyapunov function evaluation. we will use the quadratic function (4) as our Lyapunov function. The Lyapunov drift for two successive time slots has the following form.

$$\begin{aligned} &E[V(X(t+1)) - V(X(t)) \mid X(t)] \\ &= E \left[\sum_{i=1}^L X_i^2(t+1) - X_i^2(t) \mid X(t) \right] \\ &= E \left[\sum_{i=1}^L (X_i(t+1) - X_i(t))^2 \mid X(t) \right] \\ &+ 2E \left[\sum_{i=1}^L X_i(t)(X_i(t+1) - X_i(t)) \mid X(t) \right] \end{aligned} \quad (24)$$

For the the first term we have:

$$\begin{aligned} &E \left[\sum_{i=1}^L (X_i(t+1) - X_i(t))^2 \mid X(t) \right] \\ &= E \left[\sum_{i=1}^L (A_i(t+1) - \sum_{k=1}^K h_{ik}(t))^2 \mid X(t) \right] \\ &= E \left[\sum_{i=1}^L A_i^2(t+1) \mid X(t) \right] \\ &- 2E \left[\sum_{i=1}^L \sum_{k=1}^K A_i(t+1) h_{ik}(t) \mid X(t) \right] \\ &+ E \left[\sum_{i=1}^L \left(\sum_{k=1}^K h_{ik}(t) \right)^2 \mid X(t) \right] \end{aligned} \quad (25)$$

Using the the fact that $\sum_{k=1}^K h_{ik}(t) \geq 0$ we get the following inequality

$$\sum_{i=1}^L \left(\sum_{k=1}^K h_{ik}(t) \right)^2 \leq \left(\sum_{i=1}^L \sum_{k=1}^K h_{ik}(t) \right)^2 \leq K^2 \quad (26)$$

Since $\sum_{i=1}^L \sum_{k=1}^K A_i(t+1) h_{ik}(t) \geq 0$, the first term in (24) can

be bounded by

$$\begin{aligned} & E \left[\sum_{i=1}^L (X_i(t+1) - X_i(t))^2 \mid X(t) \right] \\ & \leq \sum_{i=1}^L E[A_i^2(t+1)] + K^2 \end{aligned} \quad (27)$$

Now assume that we select the servers for service according to an arbitrary order s_1, s_2, \dots, s_K . Thus, for the second term in (24) we have

$$\begin{aligned} & E \left[\sum_{i=1}^L X_i(t)(X_i(t+1) - X_i(t)) \mid X(t) \right] \\ & = E \left[\sum_{i=1}^L X_i(t)(A_i(t+1) - \sum_{k=1}^K h_{ik}(t+1)) \mid X(t) \right] \\ & = E \left[\sum_{i=1}^L X_i(t)A_i(t+1) \mid X(t) \right] \\ & - E \left[\sum_{i=1}^L \sum_{k=1}^K X_i(t)h_{is_k}(t+1) \mid X(t) \right] \end{aligned} \quad (28)$$

The first term in (28) can be written as follows.

$$E \left[\sum_{i=1}^L X_i(t)A_i(t+1) \mid X(t) \right] = \sum_{i=1}^L E[A_i(t+1)]X_i(t) \quad (29)$$

For the second term in (28) we have

$$\begin{aligned} & E \left[\sum_{i=1}^L \sum_{k=1}^K X_i(t)h_{is_k}(t+1) \mid X(t) \right] \\ & = \sum_{k=1}^K E \left[\sum_{i=1}^L X_i(t)h_{is_k}(t+1) \mid X(t) \right] \end{aligned} \quad (30)$$

Now, we introduce the following notation. We sort the queue length process at time slot t in an ascending order $X_{q_1}, X_{q_2}, \dots, X_{q_L}$, i.e. $X_{q_i}(t) \geq X_{q_{i-1}}(t)$ for all $i = 2, \dots, L$ and if $X_{q_i}(t) = X_{q_{i-1}}(t)$, then $q_i \geq q_{i-1}$. Furthermore, consider the following decomposition of the connectivity processes for each server k .

$$\begin{aligned} D_0^k &= \{G_{ik}(t+1) = 0, \text{ for all } i \in \{1, \dots, L\}\} \\ D_i^k &= \{G_{q_i k}(t+1) = 1, G_{q_{i-1} k}(t+1) = 0, \\ & \quad \text{for } i < \ell \leq L \text{ and all } i \in \{1, \dots, L\}\} \end{aligned}$$

The probability of events D_i^k is given by

$$P(D_0^k) = \prod_{i=1}^L (1 - p_{ik}), \quad P(D_i^k) = p_{q_i k} \prod_{u=i+1}^L (1 - p_{q_u k})$$

In the second term of equation (30), each term in the

summation can be rewritten as

$$\begin{aligned} & E \left[\sum_{i=1}^L X_i(t)h_{is_k}(t+1) \mid X(t) \right] \\ & = E \left[\sum_{i=1}^L X_{q_i}(t)h_{q_i s_k}(t+1) \mid X(t) \right] \\ & = \sum_{l=0}^L E \left[\sum_{i=1}^L X_{q_i}(t)h_{q_i s_k}(t+1) \mid X(t), D_l^{s_k} \right] P(D_l^{s_k}) \end{aligned} \quad (31)$$

Note that

$$E \left[\sum_{i=1}^L X_{q_i}(t)h_{q_i s_k}(t+1) \mid X(t), D_l^{s_k} \right] \geq (X_{q_l}(t) - (k-1))^+.$$

Therefore, equation (31) can be bounded by

$$\begin{aligned} & E \left[\sum_{i=1}^L X_i(t)h_{is_k}(t+1) \mid X(t) \right] \\ & \geq \sum_{l=1}^L (X_{q_l}(t) - (k-1)) p_{q_l s_k} \prod_{u=l+1}^L (1 - p_{q_u s_k}) \\ & = \sum_{l=1}^L X_{q_l}(t) p_{q_l s_k} \prod_{j=l+1}^L (1 - p_{q_j s_k}) \\ & \quad - (k-1) \sum_{l=1}^L p_{q_l s_k} \prod_{j=l+1}^L (1 - p_{q_j s_k}) \end{aligned} \quad (32)$$

For the second term in (32) we have

$$\begin{aligned} & (k-1) \sum_{l=1}^L p_{q_l s_k} \prod_{j=l+1}^L (1 - p_{q_j s_k}) \\ & = (k-1) \left(1 - \prod_{j=1}^L (1 - p_{q_j s_k}) \right) \end{aligned} \quad (33)$$

and for the first term

$$\begin{aligned} & \sum_{l=1}^L X_{q_l}(t) p_{q_l s_k} \prod_{j=l+1}^L (1 - p_{q_j s_k}) \\ & = \sum_{j=2}^L (X_{q_j}(t) - X_{q_{j-1}}(t)) \left(1 - \prod_{l=j}^L (1 - p_{q_l s_k}) \right) \\ & \quad + X_{q_1}(t) \left(1 - \prod_{j=1}^L (1 - p_{q_j s_k}) \right) \end{aligned} \quad (34)$$

Equation (29) also can be written as follows.

$$\begin{aligned} & \sum_{i=1}^L E[A_i(t+1)]X_i(t) = \sum_{l=1}^L E[A_{q_l}(t+1)]X_{q_l}(t) \\ & = \sum_{j=2}^L (X_{q_j}(t) - X_{q_{j-1}}(t)) \sum_{l=j}^L E[A_{q_l}(t+1)] \\ & \quad + X_{q_1}(t) \sum_{l=1}^L E[A_{q_l}(t+1)] \end{aligned} \quad (35)$$

Using equations (28)-(30) and (32)-(35) we have the following bound for the second term in (24).

$$\begin{aligned}
& E \left[\sum_{i=1}^L X_i(t)(X_i(t+1) - X_i(t)) \mid X(t) \right] \\
& \leq \sum_{j=2}^L (X_{q_j}(t) - X_{q_{j-1}}(t)) \sum_{l=j}^L E[A_{q_l}(t+1)] \\
& \quad + X_{q_1}(t) \sum_{l=1}^L E[A_{q_l}(t+1)] \\
& \quad - \sum_{j=2}^L (X_{q_j}(t) - X_{q_{j-1}}(t)) \sum_{k=1}^K \left(1 - \prod_{l=j}^L (1 - p_{q_l s_k}) \right) \\
& \quad - X_{q_1}(t) \sum_{k=1}^K \left(1 - \prod_{j=1}^L (1 - p_{q_j s_k}) \right) \\
& \quad + \sum_{k=1}^K (k-1) \left(1 - \prod_{j=1}^L (1 - p_{q_j s_k}) \right) \\
& \leq \sum_{j=2}^L (X_{q_j}(t) - X_{q_{j-1}}(t)) \\
& \quad \cdot \left(\sum_{l=j}^L E[A_{q_l}(t+1)] - \sum_{k=1}^K \left(1 - \prod_{l=j}^L (1 - p_{q_l s_k}) \right) \right) \\
& \quad + X_{q_1}(t) \left(\sum_{l=1}^L E[A_{q_l}(t+1)] \right. \\
& \quad \left. - \sum_{k=1}^K \left(1 - \prod_{j=1}^L (1 - p_{q_j s_k}) \right) \right) + \sum_{k=1}^K (k-1) \quad (36)
\end{aligned}$$

Now, define m as

$$m = \max_{Q \subset \{1, 2, \dots, L\}, t} \left\{ \sum_{i \in Q} E[A_i(t)] - K + \sum_{k=1}^K \prod_{i \in Q} (1 - p_{ik}) \right\}$$

Therefore equation (36) can be bounded by the following.

$$\begin{aligned}
& E \left[\sum_{i=1}^L X_i(t)(X_i(t+1) - X_i(t)) \mid X(t) \right] \\
& \leq \sum_{j=2}^L (X_{q_j}(t) - X_{q_{j-1}}(t))m + X_{q_1}(t)m + \frac{K}{2}(K-1) \\
& = X_{q_L}(t)m + \frac{K}{2}(K-1) \quad (37)
\end{aligned}$$

Putting all together and according to (24), (27) and (37) the Lyapunov drift in equation (24) is upper bounded by the following.

$$\begin{aligned}
& E[V(X(t+1)) - V(X(t)) \mid X(t)] \\
& \leq \sum_{i=1}^L E[A_i^2(t+1)] + K^2 + 2X_{s_L}(t)m \\
& \quad + K(K-1) \quad (38)
\end{aligned}$$

$$= \sum_{i=1}^L E[A_i^2(t+1)] + 2K^2 - K - \epsilon(LX_{s_L}(t)) \quad (39)$$

where $\epsilon = -\frac{2m}{L}$. According to condition (11), m is negative, therefore $\epsilon > 0$. Since $\sum_{i=1}^L X_i(t) \leq LX_{s_L}(t)$, therfore $-\epsilon(LX_{s_L}(t)) \leq -\epsilon \sum_{i=1}^L X_i(t)$. Consequently the Lyapunov drift (38) is bounded by

$$\begin{aligned}
& E[V(X(t+1)) - V(X(t)) \mid X(t)] \\
& \leq LA_{max}^2 + 2K^2 - K - \epsilon \sum_{i=1}^L X_i(t) \\
& = B - \epsilon \sum_{i=1}^L X_i(t) \quad (40)
\end{aligned}$$

in which B that has positive value is defined as

$$B = LA_{max}^2 + 2K^2 - K \quad (41)$$

Therefore, according to Lemma 2, the multi-queue multi-server system is stable under AS/LCQ as long as condition (11) is satisfied and also the time average expected congestion in the system is bounded by

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{i=1}^L E[X_i(\tau)] \leq \frac{B}{\epsilon} \quad (42)$$

which is equal to (12). \square

REFERENCES

- [1] S. Kittipiyakul and T. Javidi, "Delay-optimal server allocation in multi-queue multi-server systems with time-varying connectivities," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2319-2333, May 2009.
- [2] T. Javidi, "Rate stable resource allocation in OFDM systems: from waterfilling to queue-balancing," in *Proc. Allerton Conference on Communication, Control, and Computing* 2004.
- [3] S. Kittipiyakul and T. Javidi, "A Fresh Look at Optimal Subcarrier Allocation in OFDMA Systems," in *Proc. IEEE Conference on Decision and Control*, Dec. 2004.
- [4] S. Kittipiyakul and T. Javidi, "Resource allocation in OFDMA with time-varying channel and bursty arrivals," *IEEE Commun. Lett.*, vol. 11, no. 9, pp. 708710, Sep. 2007.
- [5] S. Asmussen, "Applied Probability and Queues." New York: Springer-Verlag, Second ed., 2003.
- [6] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, No. 12, pp. 19361949, December 1992.
- [7] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, No. 2, pp. 466478, 1993.
- [8] L. Georgiadis, M. J. Neely and L. Tassiulas, "Resource Allocation and Cross Layer Control in Wireless Networks," Now Publisher, 2006.
- [9] M. J. Neely, "Dynamic power allocation and routing for satellite and wireless networks with time varying channels," Ph.D. dissertation, Massachusetts Institute of Technology, LIDS, 2003.
- [10] P.R. Kumar and S.P. Meyn, "Stability of queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, Feb. 1995.
- [11] H. Al-Zubaidy and I. Lambadaris and I. Viniotis, "Optimal Resource Scheduling in Wireless Multi-service Systems With Random Channel Connectivity," (to appear in) *Proceeding of IEEE Global Communications Conference (GLOBECOM 2009)*, Nov. 2009.
- [12] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "On Achieving 100% throughput in an input-queued switch," *IEEE Transactions on Communications*, vol. 47, pp. 12601272, August 1999.
- [13] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," *IEEE Journal on Selected Areas in Communications, Special Issue on Wireless Ad-hoc Networks*, vol. 23, No. 1, pp. 89103, January 2005.

- [14] E. Leonardi, M. Melia, F. Neri, and M. Ajmone Marson, "Bounds on average delays and queue size averages and variances in input-queued cell-based switches," *Proceedings of IEEE INFOCOM*, 2001.
- [15] A. Ganti, E. Modiano and J. N. Tsitsiklis, "Optimal transmission scheduling in symmetric communication models with intermittent connectivity," *IEEE Transactions on Information Theory*, vol. 53, Issue 3, March 2007.